ELSEVIER

Gene 179 (1996) 257–262

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

Isolation and sequencing of the gene encoding Sp23, a structural protein of spermatophore of the mealworm beetle, *Tenebrio molitor*

Xu Feng *, George M. Happ

Department of Biology, University of Vermont, Burlington, VT05405-0086, USA

Received 2 November 1995; revised 28 March 1996; accepted 7 May 1996

Abstract

The cDNA for Sp23, a structural protein of the spermatophore of *Tenebrio molitor*, had been previously cloned and characterized (Paesen, G.C., Schwartz, M.B., Peferoen, M., Weyda, F. and Happ, G.M. (1992a) Amino acid sequence of Sp23, a structure protein of the spermatophore of the mealworm beetle, *Tenebrio molitor*. J. Biol. Chem. 257, 18852–18857). Using the labeled cDNA for Sp23 as a probe to screen a library of genomic DNA from *Tenebrio molitor*, we isolated a genomic clone for Sp23. A 5373-base pair (bp) restriction fragment containing the *Sp23* gene was sequenced. The coding region is separated by a 55-bp intron which is located close to the translation start site. Three putative ecdysone response elements (EcRE) are identified in the 5′ flanking region of the *Sp23* gene. Comparison of the flanking regions of the *Sp23* gene with those of the *D-protein* gene expressed in the accessory glands of *Tenebrio* reveals similar sequences present in the flanking regions of the two genes. The genomic organization of the coding region of the *Sp23* gene shares similarities with that of the *D-protein* gene, three *Drosophila* accessory gland genes and two *Drosophila* 20-OH ecdysone-responsive genes.

Keywords: Ecdysone response element; Insect; Nucleotide sequence; Male accessory gland; Spermatophorin

1. Introduction

The accessory glands of male insects play diverse roles in insect reproduction (Leopold, 1976; Happ, 1992). The yellow mealworm beetle, *Tenebrio molitor*, has two pairs of such glands: the smaller tubular accessory glands (TAGs) and the larger bean-shaped accessory glands (BAGs) (Gadzama, 1972; Gerber, 1976). Products of both pairs of glands form the spermatophore that packages the sperm for transfer from male to female (Gadzama and Happ, 1974). The TAGs secrete a proteinaceous fluid which is mixed with the sperm within the lumen of the spermatophore (Happ et al., 1977; Black et al., 1982). The BAGs secrete sparingly soluble proteins, most of which are precursors of the structural proteins, known as spermatophorins (Happ, 1987; Shinbo et al., 1987). These proteins form the walls of the spermatophore (Gadzama and Happ, 1974). One of the proteins produced by the BAGs, an intriguing soluble trehalase, has been localized on the surface of the spermatophore with a specific polyclonal antibody and the cDNA for the trehalase has been cloned and characterized (Takiguchi et al., 1992; Yaginuma et al., 1996). Another protein produced by the BAGs, a 23-kDa spermatophorin designated Sp23, is a structural component that was mapped to a layer within the wall of the spermatophore with a monoclonal antibody (Shinbo et al., 1987). The cDNA for Sp23 has been isolated from a cDNA expression library produced from the adult BAGs of *Tenebrio molitor* (Paesen et al., 1992a). As deduced from the nucleotide (nt) sequence of the cDNA for Sp23, the Sp23 protein is composed of 175 amino acids (aa). In this paper, we report the isolation and sequencing of a genomic clone for Sp23.

* Corresponding author. Present address: Department of Pathology, Jewish Hospital at Washington University Medical Center, 216 South Kingshighway, St. Louis, MO 63110, USA. Tel. +1 314 4545084; Fax: +1 314 4545505; e-mail: fengx@medicine.wustl.edu

Abbreviations: aa, amino acid(s); BAGs, bean-shaped accessory glands; bp, base pair(s); cDNA, DNA complementary to RNA; EcRE, ecdysone response element(s); GCG, Genetics Computer Group (Madison, WI, USA); kb, kilobase(s) or 1000 bp; nt, nucleotide(s); ORF, open reading frame; TAGs, tubular accessory glands.

PII S0378-1119(96)00372-1

2. Experimental and discussion

2.1. Isolation and sequencing of a genomic clone of Sp23 gene

Genomic clones of *Sp23* gene were isolated from the genomic DNA library of *Tenebrio molitor* (Paesen et al., 1996) using a probe made from a cDNA clone for Sp23 (Paesen et al., 1992a). After screening about 800 000 plaques, 12 positive plaques were obtained from the genomic DNA library. One positive plaque was randomly selected for further analysis. Restriction digestion of the phage DNA from the positive clone with *Bam*HI and *Xba*I generated three fragments in addition to the phage arms: a 4-kilobase (kb) fragment, a 4.5-kb fragment and a 5.4-kb fragment. Only the 5.4-kb fragment positively reacted with the *Sp23* cDNA probe on a southern blot. The 5.4-kb fragment was subcloned into a pBluescript plasmid and, subsequently, sequenced. Figure 1 shows the nt sequence of part of the 5.4-kb fragment (from nt 1 to nt 3060) and the deduced aa sequence. The coding sequence from the genomic clone is very similar, but not identical, to that of the reported cDNA for Sp23 (Paesen et al., 1992a). There are eight nt differences, resulting in three changes in aa residues. The differences may indicate that the allelic variants exist in the *Sp23* gene, and that the genomic clone and the cDNA clone are from different individuals. Consistently, the southern analysis of the *Sp23* gene suggested that different alleles exist in the *Sp23* gene (Paesen et al., 1992a). The open reading frame (ORF) contains one translation start codon (2307–2309), which fits the pattern of a consensus sequence for a translation start site [(A/G)NNATGG] (Kozak, 1983). A putative TATA-box, found at position 2234–2239, agrees with the consensus sequence (Buchei and Trifonov, 1986). The coding region is separated by a 55-bp intron (2326–2381). This intron follows the GT-AG rule and is pyrimidine-rich at its 3′ terminus (Breathnach and Chambon, 1981). The polyadenylation signal sequence (AATAAA) is located at positions 2954–2959.

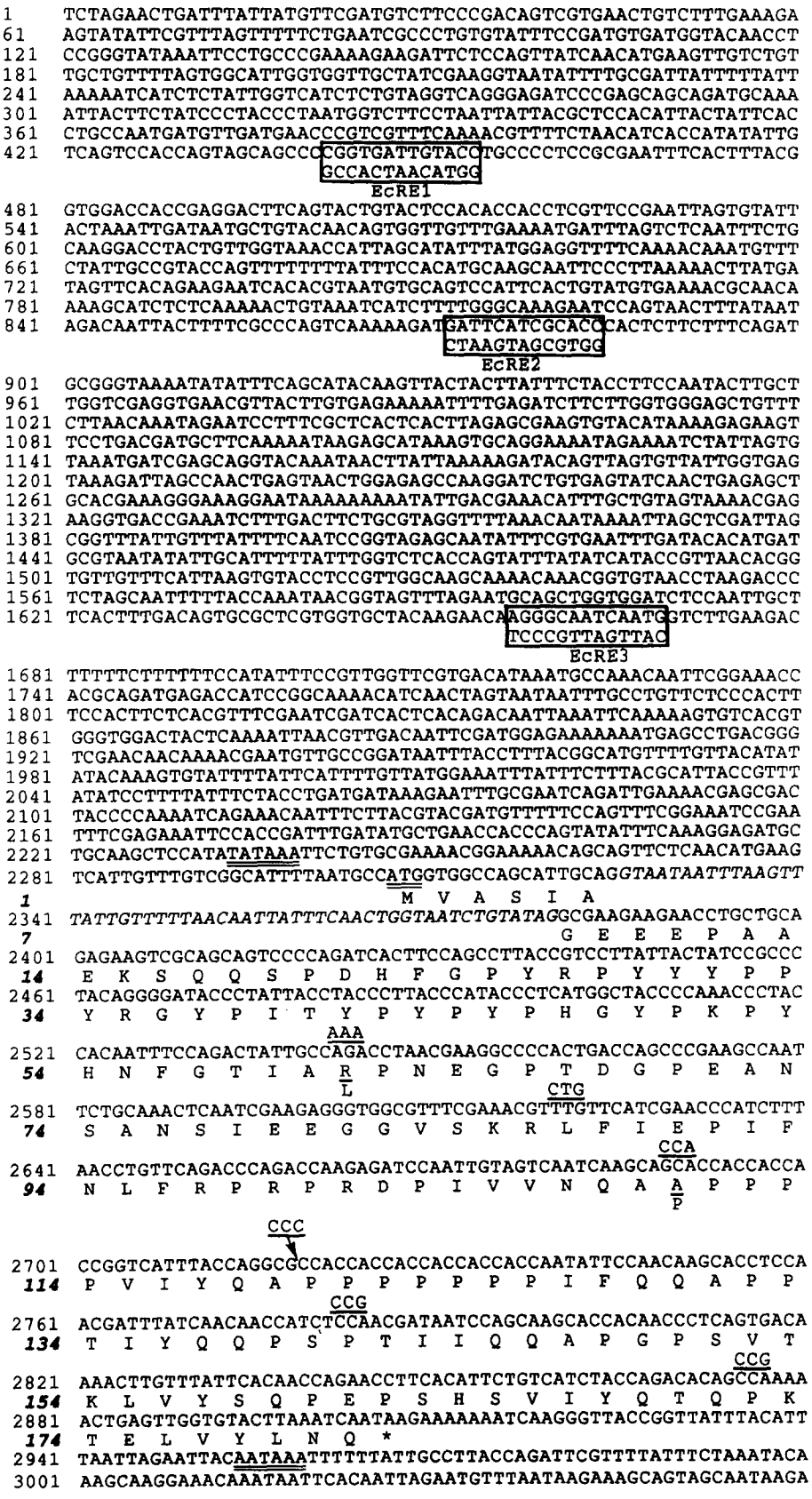
2.2. Putative EcRE in the upstream region of the Sp23 gene

Analogy with trehalase, another secretory protein of the BAG (Yaginuma and Happ, 1989), suggests that the molting hormone, 20-OH ecdysone, is required for expression of the *Sp23* gene. Thus, we searched for putative EcRE in the regions flanking the *Sp23* gene. The EcRE consensus sequence, [PuG(G/T)T(C/G) A(N)TG(C/A)(C/A)(C/t)Py] described by Antoniewski et al. (1993) in *Drosophila melanogaster*, was used to search for putative EcRE with the Findpattern program of GCG (Genetics Computer Group). No sequence identical to the fly consensus sequence was found in the upstream or downstream regions of this beetle gene. However, when the constraints were relaxed so that two mismatches were permitted, four possible EcRE were discovered in the upstream region and two putative EcRE in downstream region of the *Sp23* gene. As a general rule, the first two nt (TG) of the second half site are highly conserved among the other EcRE as well as in vertebrate steroid hormone response elements (Beato, 1989; Antoniewski et al., 1993). Since three of these six putative EcRE have a variant nt at the first two positions of the second half site, they are considered unlikely to function as EcRE, and excluded. The remaining three putative EcRE are shown in Fig. 1. EcRE 1, 5′-CGGTGATTGTACC-3′, is located at nt 442–454. EcRE 2, 5′-GGTGCGATGAATC-3′, is located at nt 872–884 on the complementary strand. EcRE 3, 5′-CATTGATTGCCCT-3′, is located at nt 1657-1669 on the complementary strand.

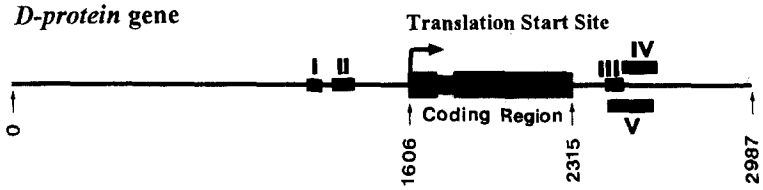
2.3. Comparison of flanking regions of the Sp23 gene with those of the D-protein gene

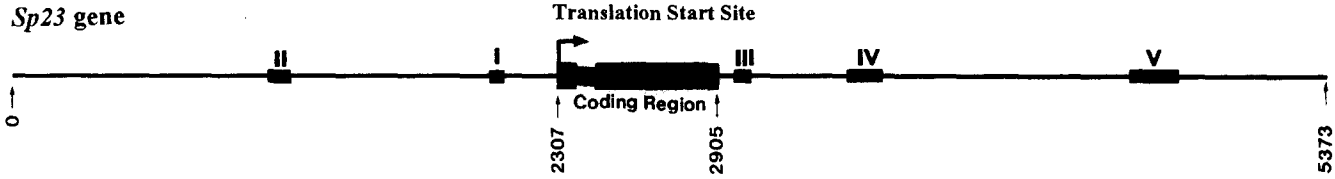
Like the *Sp23* gene expressed in the BAGs, the *D-protein* gene, which codes for the D-protein in the TAGs, is expressed at the beginning of the adult stage, about 4 days after the pupal peak of ecdysteroids (Paesen et al., 1992a,b). The parallel patterns of expression of the two genes may indicate that they are transcriptionally regulated by similar mechanisms. The nt sequence of a genomic clone for the D-protein has been reported elsewhere (Paesen et al., 1996). The Bestfit program of GCG was used to compare the flanking regions of the *Sp23* and *D-protein* genes. In the 5′ flanking regions, two segments of similarity between the two genes were discovered. The first one (Segment I) is 20 bp long with 80% similarity between the two genes (Fig. 2). The second segment (Segment II) is 56 bp long with 69.8% similarity between the two genes. In the 3′ flanking regions of the *Sp23* and *D-protein* genes, there are three interesting regions of similar sequence in the two genes.

Fig. 1. Genomic nt sequence of the *Sp23* gene and its flanking regions (GenBank accession No. U39658). Only the first 3060 sequenced nucleotides are shown here. The putative EcREs are numbered and indicated by boxes. The putative TATA-box, the translation start site and the polyadenylation signal are double underlined, respectively. The intron is in italics. The stop codon is indicated by an asterisk. The differences between the genomic sequence and the cDNA sequence (Paesen et al., 1992a) are indicated. The triplets overlined in the genomic sequence are different from those of the cDNA. Their cDNA counterparts are provided above the overlines. Two of the cDNA counterparts encode different amino acids. The amino acids encoded by these two cDNA counterparts are provided under the corresponding amino acids. One triplet (CCC) is missing in the genomic sequence. The location of the missing triplet is indicated by a small arrow below the triplet (between 2718 and 2719).


EcRE1
EcRE2
EcRE3

A





B

Segment I

```
D   TAGATCTGGCAATTTATTTC
    | | | ||| ||||||||||
S   TTGTTATGGAAATTTATTTC
```

Segment III

```
D   AATTTTAAAGACTGAGAAAAGAAAATTATAAT
    | || |||| || || || ||||  |||||
S   ATTTCTAAATACAAAGCAAGGAAACAAATAAT
```

Segment II

```
D   AAATTTTTTTGTTGCCAATGATTGAGTAAACAAGTCAAAGACGTTATTAAAAAGTT
    ||| | | ||  ||  |||||| ||| |   |   |||| |  |||||||||||
S   AAAATCTATTAGTGTAAATGATCGAGCAGGTA...CAAATAACTTATTAAAAAGAT
```

Segment IV

```
D   TATAATGGCTATAATTCAACAGGATGAAAATTCCAACACGAGCCGTAGGCGAGTGGTGGAA...
    |||||| ||||| |||||   | | |  |||||| |||  | |  |||  ||  |||||
S   TATAAT.GCTATTATTCACTCGCACG...ATTCCACCACTCGACTACGGCTCGTCTTGGAATTT

    TCATCCGAGTGAATAATAGCCAT..........CATTGAAGAATGTACTACT
    ||||||  || ||| ||||||||          |||||||||||||||| |
    TCATCCTCGTAAATTATAGCCATAATTATACGCTCATTGAAGAATGTACTATT
```

Segment V

```
D   CAAATTTTAAAGACTGAGAAAAGAAAATTATAATGGCTATAATTCAACAGGATGAAAATTCCAA
    | |||  || |  ||   |  ||   || || ||||||||||||| | ||||||||||||||||
S   CGAATAGTACATTCTTCAATGAGCGTATAATTATGGCTATAATTTACGAGGATGAAAATTCCAA

    CACGAGCCGTAGGCGAGTGGTGGAATCATCCGAGTGAATAATAGCCATCA
     |||||||||| |||||||||||||| | |||||||||||||| ||| |
    GACGAGCCGTAGTCGAGTGGTGGAATCGTGCGAGTGAATAATAG.CATTA
```

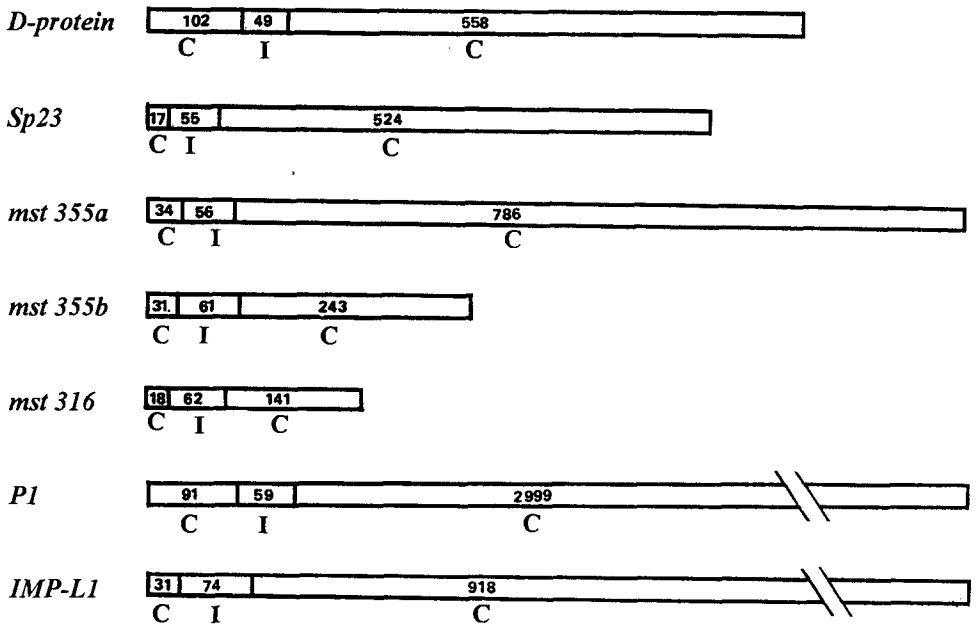


Fig. 3. Comparison of the genomic organization of the coding regions of the *D-protein* and *Sp23* genes with those of *mst 355a*, *mst 355b*, *mst 316*, *P1* and *IMP-L1*. C, coding region; I, intron. The numbers inside the boxes indicate the length of the introns or the coding regions.

The first one, designated Segment III, is 32 bp long with 68.7% similarity between the two genes (Fig. 2). The second one, designated Segment IV, is 117 bp long with 77.7% similarity between the two genes. The third one, designated Segment V, is 114 bp long with 78% similarity between the two genes. The existence of the similar sequences in the flanking regions of these two genes is intriguing. However, the potential roles of any of these similar sequences in transcriptional regulation of expression of these two genes remain purely speculative without a functional assay.

2.4. *Comparison of the genomic organization of the* Sp23 *gene with genes in other insects*

In their overall organization, the *D-protein* and *Sp23* genes share similarities with other insect genes, notably three *Drosophila* genes (*mst 355a*, *mst 355b* and *mst 316*) expressed in the male accessory glands (Monsma and Wolfner, 1988; DiBenedetto et al., 1990), a *Drosophila* gene (*P1*) expressed in fat body cells (Maschat et al., 1990), and a *Drosophila* gene (*IMP-L1*) expressed during imaginal disc morphogenesis (Natzle et al., 1992). The coding regions of all these genes are separated by small introns which are located close to the translation start sites (Fig. 3). The biological significance of the similar genomic organization of the coding regions among these genes remains unclear.

Acknowledgement

We thank Guido C. Paesen for advice and assistance during the course of this study. This work was supported in part by a NIH grant AI-15662 (to G.M.H.) and a Grant-in-Aid of Research from Sigma Xi (to X.F.).

References

Antoniewski, C., Larval, M. and Lepesant, J.-A. (1993) Structural features critical to the activity of an ecdysone receptor binding site. Insect Biochem. Mol. Biol. 23, 105–114.

Beato M. (1989) Gene regulation by steroid hormones. Cell 56, 335–344.

Black, P.N., Landers, M.H. and Happ, G.M. (1982) Cytodifferentiation in the accessory glands of *Tenebrio molitor*. VIII. Crossed immunoelectrophoretic analysis of terminal differentiation in the postecdysial tubular accessory glands. Dev. Biol. 94, 106–115.

Breathnach, R. and Chambon, P. (1981) Organization and expression

Fig. 2. Similar sequences present in the 5′ and 3′ flanking regions of the *D-protein* and *Sp23* genes. (A) Schematic locations of these sequences (they are designated Segments I, II, III, IV and V) in the two genes. (B) Nucleotide sequence comparison of these fragments between the *Sp23* and *D-protein* genes. D, *D-protein* gene; S, *Sp23* gene. Segment I is located at positions 1264–1283 in the *D-protein* gene, and at positions 2004–2023 in the *Sp23* gene; Segment II is located at positions 1430–1485 in the *D-protein* gene, and at positions 1127–1179 in the *Sp23* gene; Segment III is located at positions 2569–2600 in the *D-protein* gene, and at positions 2988–3019 in the *Sp23* gene; Segment IV is located at positions 2595–2697 in the *D-protein* gene, and at positions 3541–3653 in the *Sp23* gene; Segment V is located at positions 2570–2680 in the *D-protein* gene, and at positions 4719–4832 in the *Sp23* gene.

of eukaryotic split genes coding for proteins. Annu. Rev. Biochem. 50, 349–383.

Buchei, P. and Trifonov, E.N. (1986) Composition and analysis of eukaryotic POL II promoter sequences. Nucleic Acid Res. 14, 10009–10026.

DiBenedetto, A.J., Harada, H.A. and Wolfner, M.F.: Structure, cell-specific expression, and mating-induced regulation of a *Drosophila melanogaster* male accessory gland gene. Dev. Biol. 139, 134–148.

Gadzama, N.M. (1972) A histological, histochemical and electron microscopical study of the spermatophore and internal reproductive organs of *Tenebrio molitor* L. (Coleoptera, Tenebrionidae). Ph.D. thesis, New York University, New York.

Gadzama, N.M. and Happ, G.M. (1974) The structure and evacuation of the spermatophore of *Tenebrio molitor* L. (Coleoptera: Tenebrionidae). Tissue Cell 6, 211–221.

Gerber, G.H. (1976) Reproductive behaviour and physiology of *Tenebrio molitor* (Coleoptera: Tenebrionidae). III. Histogenetic changes in the internal genitalia, mesenteron, and cuticle during sexual maturation. Can. J. Zool. 54, 990–1002.

Happ, G.M., Yuncker, C. and Huffmire, S.A. (1977) Cytodifferentiation in the accessory glands of *Tenebrio molitor*. II. Patterns of leucine incorporation in the tubular glands of post-ecdysial adult males. J. Exp. Zool. 200, 223–236.

Happ, G.M. (1987) Accessory gland development in mealworm beetles. In: Law, J. (Ed.), Molecular Entomology. Alan R. Liss, New York, pp. 433–442.

Happ, G.M. (1992) Maturation of the male reproductive system and its endocrine regulation. Annu. Rev. Entomol. 37, 303–320.

Kozak, M. (1983) Comparison of initiation of protein synthesis in procaryotes, eucaryotes and organelles. Microbiol. Rev. 47, 1–45.

Leopold, R.A. (1976) The role of male accessory glands in insect reproduction. Annu. Rev. Entomol. 21, 199–221.

Maschat, F., Dubertret, M.-L., Therond, P., Claverie, J.-M. and Lepesant, J.-A. (1990) Structure of the ecdysone-inducible *P1* gene of *Drosophila melanogaster*. J. Mol. Biol. 214, 359–372.

Monsma, S.A. and Wolfner, M.F. (1988) Structure and expression of a *Drosophila* male accessory gland gene whose product resembles a peptide pheromone precursor. Genes Dev. 2, 1063–1073.

Natzle, J.E., Robertson, J.P., Majumdar, A., Vesenka, G.D., Enlow, B. and Clark, K.E. (1992) Sequence and expression of *IMP-L1*, an ecdysone-inducible gene expressed during *Drosophila* imaginal disc morphogenesis. Dev. Genet. 13, 331–344.

Paesen, G.C., Schwartz, M.B., Peferoen, M., Weyda, F. and Happ, G.M. (1992a) Amino acid sequence of Sp23, a structure protein of the spermatophore of the mealworm beetle, *Tenebrio molitor*. J. Biol. Chem. 257, 18852–18857.

Paesen, G.C., Weyda, F. and Happ, G.M. (1992b) Deduced sequences show multiple repeats in two D proteins from the tubular accessory glands of *Tenebrio molitor*. Insect Biochem. Mol. Biol. 22, 387–398.

Paesen, G.C., Feng, X. and Happ, G.M. (1996) Structure of a D-protein gene and amino acid sequence of the highly-repetitive D-proteins secreted by the accessory glands of the mealworm beetle. Biochim. Biophys. Acta (in press).

Shinbo, H., Yaginuma, T. and Happ, G.M. (1987) Purification and characterization of a proline-rich secretory protein that is a precursor to a structural protein of an insect spermatophore. J. Biol. Chem. 262, 4794–4799.

Takiguchi, M., Niimi, T., Su, Z.-H. and Yaginuma, T. (1992) Trehalase from male accessory gland of an insect, *Tenebrio molitor*. cDNA sequencing and developmental profile of gene expression. Biochem. J. 288, 19–22.

Yaginuma, T. and Happ, G.M. (1989) 20-Hydroxyecdysone acts in the male pupa to commit accessory glands toward trehalase production in the adult mealworm beetle (*Tenebrio molitor*). Gen. Comp. Endocrinol. 73, 173–185.

Yaginuma, T., Mizuno, T., Mizuno, C., Ikeda, M., Wada, T., Hattori, K., Yamashita, O. and Happ, G.M. (1996) Trehalase incorporated into the wall of the spermatophore from the bean-shaped accessory gland of the male mealworm beetle, *Tenebrio molitor*. Purification, kinetic properties, and localization. J. Comp. Physiol. B 166, 1–10.